图书流通数据分析的H类指数遴选及应用*

——以西北农林科技大学图书馆为例

■ 净玲娣 王立宏 王芹 李雅
西北农林科技大学图书馆 杨凌 712100

摘 要:[目的/意义]优化藏书采访工作和流通管理,实现对藏书采访、藏书总量和结构等状况适时监测、评估、调控等的标准化管理。[方法/过程]通过对H类指数中的H指数、A指数和G指数相关性、变异性和适应性分析,遴选适用于图书流通数据分析的A指数,并根据帕累托定律采用A指数对西北农林科技大学图书馆2014－2018年的前20%的图书流通数据分别进行统计分析。[结果/结论]研究结果表明:平均A指数对中图法一级类目与读者性别、读者专业、读者年级以及图书出版社、图书出版年、单位等数据统计分析具有较理想的表现,能够更全面、准确地遴选核心读者群、核心出版者和核心出版年;同时发现,布拉德福定律不仅能够应用于期刊管理,而且能够应用于藏书管理。本文研究能够为探索H类指数分析高校图书馆藏书流通数据提供参考依据。

关键词:H类指数 平均A指数 变异系数 相关性 图书流通数据
分类号:G252
DOI:10.13266/j.issn.0252－3116.2020.09.008

1 引言

H指数是由美国学者J. E. Hirsch于2005年提出的旨在评价个人成就的一种指标,并将其定义为:当且仅当一个科学家有H篇引文数至少为H的论文,同时剩余论文的引文数都小于H时,该科学家的成就分值为H[1]。其主要优点是:计算简单,能同时进行数量测度和质量测度,且增长比较缓慢,在自然状态下不太可能突然增长很多,这保证了评价结果在一定时期内的有效性,而且当个别数据出现问题时对H指数影响很小。周志峰[2]进行了H指数高校图书馆借阅数据可行性及应用分析,认为图书借阅H指数能够更加客观地反映图书利用和读者借阅情况,能够克服传统的绝对数据(如借阅频次),相对数据(如平均借阅量)无法在藏书集合(读者群)之间进行客观比较,无法相对公正的衡量和确定核心藏书群和读者群等缺点。但H指数随着时间变化只升不降,同时极值数据影响力反映不足,因此它对数据变化不够灵敏[3],也不能用于跨领域比较[4],为此,L. Egghe在H指数基础上提出G指数[5-6]。金碧辉和R. Rousseau建议将H核心集中论文的引用平均数作为一个测度指标,称之为A指数[3]。这样即使H指数不增长,A指数也可反映H核内的变动,因此A指数立即成为一个热点问题。

刘玉仙和R. Rousseau[7]将H指数应用于图书馆借阅量分析;梁晓红[8]认为纸质馆藏资源具有数字馆藏资源不可替代的优势,如何发挥纸质馆藏潜能和提高经费产出效能,是高校图书馆面临的巨大挑战;周春雷[9]从累计利用率、累计借阅量及H指数3个角度分析读者需求,为图书馆馆藏纸质资源的利用提供了分析依据;孔超和丁璇[10]基于H指数、G指数借阅数据可行性分析,认为H指数能够更加准确的评价馆藏图书质量;钱令飞和汪荣[11]认为读者借阅数据和藏书借阅数据序列类似于H指数原始定义中论文数与引文数的数据序列,从而可以计算其H指数;张晓阳、方友亮和宋新平[12]在理论上提出许多基于H指数的衍生指数,在某些方面填补了H指数的不足。上述研究从不同侧面论述了H类指数的应用,但都是从逻辑推理或者直接基于H指数的A指数、G指数研究,缺乏H

* 本文系2019年西北农林科技大学本科教育教学改革研究项目"图书馆藏书标准化管理建设研究"(项目编号:JY1903221)研究成果之一。
作者简介:净玲娣(ORCID:0000－0003－4089－3280),馆员;王立宏(ORCID:0000－0002－1705－8626),馆员;王芹,采编部主任,副研究馆员;李雅(ORCID:0000－0002－8898－5689),研究馆员,通讯作者,E－mail:liyachenggong@126.com。
收稿日期:2019－06－18 修回日期:2019－11－01 本文起止页码:65－72 本文责任编辑:王传清

类指数的统计分析、遴选以及藏书流通数据的实证分析，本研究通过 R 软件将中图法一级类目与 Excel 中各指标列变量相关性进行卡埃方检验，采用 LEFT 函数法提取数据，并采取 H、A、G 指数两两相关性分析法及其变异系数法及遴选 A 指数，对西北农林科技大学图书馆 2014－2018 年图书流通数据进行实证分析，为高校图书馆有效分析藏书流通数据，确定核心读者群、核心藏书群、核心藏书群、核心出版者以及核心出版年等提供新方法。

2 数据来源与获取

2.1 研究对象的确定

根据帕累托定律[13]选取西北农林科技大学（以下简称"我校"）图书馆 2014－2018 年的前 20% 的流通借阅数据作为研究对象，对图书流通的 7 个方面影响因素进行统计分析，对通过卡埃方检验具有显著关联性的数据进行 H 指数、A 指数和 G 指数计算与分析。图书流通数据提取时间为 2019 年 3 月 16 日。

2.2 数据来源与方法

第一，将从图书管理系统中导出的借阅数据，通过 Excel 左截取 LEFT 函数分别截取索书号的前 1、2 或 3 位字符。分别以《中国图书馆分类法》（以下简称"中图法"）一、二级类目的索书号为行变量，以复本量、书名、著者、出版者、出版年、ISBN 号、读者证件号、姓名、性别、单位、年级、读者类别等为列变量分别形成 2014－2018 年的前 20% 的图书流通数据：149 526、119 801、105 631、98 413、112 795条数据矩阵，并进行数据规范化清理。

第二，通过 R 软件将中图法一级类目与各影响因素进行卡埃方检验，以筛选图书流通数据内相互关联的指标，将欲计算的数据整理后保存于 Excel 文件中，并通过 Visual C#编程将文件中数据赋值于二维数组（x1，y1，x1，y2，…，xi，yj，…，xm，yn）中，生成矩阵数列，进而按照 3 个指数的定义及参考文献[2]，分别计算相应的行变量和列变量的 H、A、G 指数。

第三，计算 2014－2018 年各影响因素列联表 H 指数、A 指数和 G 指数平均数的两两相关性，选择相关性强的指数中相对易获取、区分度高者进行数据分析。

第四，变异系数是反映总体标志值的单位差异程度或离散程度的指标。在多指标评价中如果某项指标对所有该评价对象观测值的变异程度较大，说明该指标具有明确区分各被评价对象的能力，应赋予较大权重[14]。因此笔者选用变异系数法进一步筛选评价 H 类指数。

第五，通过相关分析、变异系数法筛选 H 类指数，利用筛选出的 A 指数平均值确定核心读者群、核心藏书群、核心出版者以及核心出版年。

3 数据处理与分析

3.1 H 类指数计算与筛选

利用 R 软件对 2014－2018 年的 5 年的图书流通数据的 7 个指标（影响因素）分别与其中图法一级类目进行相关性分析，发现简单相关检验中均有 p_value < 0.05，表明藏书流通 7 个指标与其中图法一级类目均具有高度相关性。

3.1.1 图书流通数据 H 类指数相关性分析

应用线性相关性评判 H 类指数间相关性通常的判断标准是：$|r|<0.3$ 称为微弱相关，$0.3<|r|<0.5$ 称为低度相关，$0.5<|r|<0.8$ 称为显著相关，$0.8<|r|<1$ 称为高度相关；当 $r>0$ 时，表示 x 与 y 为正相关，当 $r<0$ 时，表示 x 与 y 为负相关[15]。如表 1 所示：

表 1 2014－2018 年 H 类指数相关性

类 H 指数相关性	相关性分析	一级类目_读者单位	一级类目_读者专业	一级类目_出版者	一级类目_出版年	一级类目_读者性别	一级类目_读者年级	一级类目_读者类别
H 与 A	相关系数	0.828 7	0.815 7	0.445 4	0.962 1	0.997 2	0.796 9	0.838 8
	显著性	p-value = 5.6871e-07	p-value = 0.025 3	p-value < 2.2220e-16	p-value = 0.002 1	p-value = 1.1901e-05	p-value = 1.0173e-08	p-value = 0.075 8
H 与 G	相关系数	0.722 7	0.911 9	0.850 8	0.997 8	0.971 7	0.941 5	0.393 3
	显著性	p-value = 6.6491e-05	p-value = 0.004 22	p-value < 2.2032e-16	p-value = 7.0603e-06	p-value = 0.001 2	p-value < 2.2e-16	p-value = 0.512 5
G 与 A	相关系数	0.573 3	0.537 8	0.641 2	0.969 9	0.957 0	0.577 3	-0.135 3
	显著性	p-value = 0.003 4	p-value = 0.021 3	p-value < 2.2350e-16	p-value = 0.001 35	p-value = 0.002 7	p-value = 0.000 3	p-value = 0.828 3

（1）H 指数与 A 指数相关性分析：一级类目分别与读者单位、读者专业、读者性别、读者学历以及出版年、出版者的 H 指数与 A 指数的相关系数均 >0.8，单相关检验中均有 p_value <0.05，即这些影响因素的 H 指数与 A 指数高度相关；一级类目与读者类别的 H 指数与 A 指数相关系数 >0.8，H_0 不相关性假设 0.076 > 0.05，接受 H_0 假设，即 H 指数与 A 指数不相关。

（2）H 指数与 G 指数相关性分析：一级类目分别与读者专业、读者性别、读者学历、读者单位、出版者、出版年 H 指数与 G 指数相关系数均 >0.72，且均有 p-value <0.05，呈高度相关性或显著相关；而一级类目－读者类别相关相关系数为 0.39，0.39 <0.4，p-value = 0.512 5，接受 H_0 不相关假设，一级类目－读者类别 H-G 指数相关性不显著。

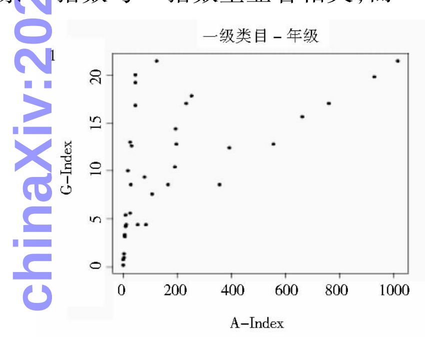
（3）G 指数与 A 指数相关性分析：一级类目分别与出版年、读者性别 G 指数与 A 指数相关系数 > 0.95，p-value <0.05，即一级类目与出版年、读者性别 G 指数与 A 指数呈高度相关；一级类目分别与读者单位、读者学历、读者专业、出版者 G 指数与 A 指数的相关系数位于 0.5－0.8 之间，p-value <0.05，即上述因素 G 指数与 A 指数呈显著相关；而一级类目－读者类别的 G 指数与 A 指数 p-value = 0.828 3 且 >0.05，接受 H_0 不相关假设，即一级类目－读者类别 G-A 指数不相关。

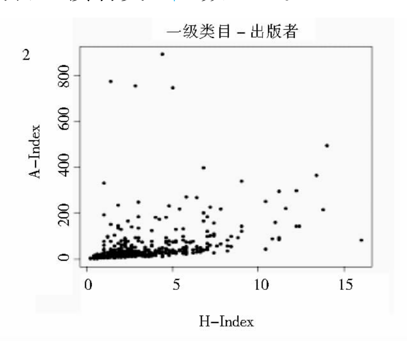
综上所述，除了一级类目－读者类别 H、A、G 指数两两不相关外，其余一级类目分别与读者专业、读者性别、读者学历、读者单位、出版者、出版年的 H 指数与 A 指数、H 指数与 G 指数、G 指数与 A 指数均呈高度相关或显著相关，均有 p-value <0.05，说明它们所提供的信息具有一致性，可通过变异系数法遴选区分度高的 A 指数进行核心读者群、核心出版者与核心出版年等分析。

虽然从各影响因素的 H 类指数之间相关性和区分度可以进行 H 类指数的筛选，但是其中是否含有具体适用条件还需进一步分析。

3.1.2　图书流通数据 H 类指数相关性图解分析

一级类目－年级 A-G 指数虽然相关系数为 0.577 3，p-value< 0.000 3，呈显著相关，但是通过对图 1－1 作进一步分析，可见只有在 A 指数接近 400 以后，才显示一级类目－学历年级 A-G 指数显著相关性；虽然一级类目－出版者 A-H 指数相关度呈高度相关，但是图 1－2 显示，H 指数与 A 指数相关性体现于 H 指数 >10。





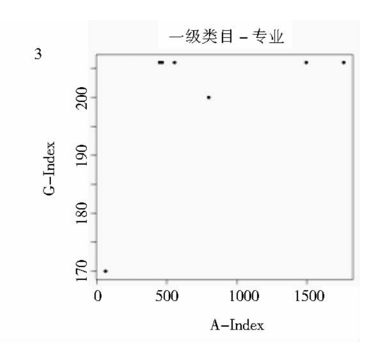


图 1　图书流通数据 H 类指数相关性特征性

综合分析表 1 与图 1，一级类目分别与出版年、出版者、读者单位、读者性别、读者性别、读者专业的 H、A、G 指数图相关性一致，其中一级类目－出版年的 H-G 指数、一级类目－年级的 A-G 指数呈有条件的相关；一级类目－类别的 H、A、G 指数均不相关。

3.1.3　2014－2018 年流通数据 H 类指数筛选

分别计算一级类目与单位、年级、读者类别、性别、专业、出版者和出版年 H 指数、A 指数和 G 指数及其平均数，在具有相关性的 H 类指数中通过区分度筛选最适合的指数，在计算过程中使用 2014－2018 年藏书流通数据平均 H 类指数计算变异系数来进一步筛选评价指数，计算变异系数公式为：

$$v_i = \sigma_i / \overline{x_i} \quad (i = 1, 2, \cdots, n) \qquad \text{公式(1)}$$

公式(1)中，σ_i 表示第 i 项指标的标准差；$\overline{x_i}$ 表示第 i 项指标的平均数，n 表示指标的个数[16]。

H 类指数变异系数越大，其区分度越好[17]，由表 2 可知，2014－2018 年 7 个影响因素的 3 个指数中，A 指数的变异系数最大，分别为 0.496 6、0. 766 1、2.021 8、1.772 7、0.070 0、1.438 0、1.226 3，H 指数次之，G 指数最小，可见 H 指数、A 指数、G 指数变异系数大小与其总体平均值跨度呈正相关，总体平均值跨度越大，区分度越好[18]。除了一级类目－读者类别外，其他 6 个指标的 H 指数与 A 指数、G 指数均显著或高度相关。A 指数的优点基于 H 指数，又考虑藏书流通

表 2 H 指数、A 指数和 G 指数分布对比

H 类指数	H 指数			A 指数			G 指数		
	年度分布	5 年均值	变异系数	年度分布	5 年均值	变异系数	年度分布	5 年均值	变异系数
类目_读者单位	14－15.5	14.20	0.223 0	311－382.7	324.6	0.496 6	20.3－21.0	20.5	0.130 0
类目_读者专业	8.2－19.4	15.8	0.235 9	846－1 037.6	797.4	0.766 1	17－20.6	20	0.067 1
类目_出版者	2.5－2.9	2.6	1.003 1	33.3－55.3	43.2	2.021 8	6.5－8.6	7.3	0.890 9
类目_出版年	5.8－6.4	6.1	0.985 6	113.7－163.3	131.9	1.772 7	4.7－10.5	9.1	0.826 9
类目_读者性别	16－20.5	19.4	0.030 9	2 490－3 600.8	2 881.2	0.0549	20.6－20.6	20.6	0
类目_读者学历	0.2－18.8	7.2	0.687 7	0.4－1013.9	190	1.438 0	0.2－21.4	10.7	0.624 1
类目_读者类别	13.6－20.8	17.1	0.174 6	162－235.1	190	1.226 3	21.2－21.4	21.26	0.004 2

数据极值的贡献，A 指数按年度计算同时克服了 H 指数的只升不降的缺点，同时变异系数最大，区分度最好；G 指数虽然考虑了极值的贡献，但计算繁琐，区分度较小，因此选择 A 指数作为一级类目分别与上述读者分析、出版者、出版年矩阵数据的评价指数；而一级类目－读者类别 H 指数、A 指数、G 指数不相关，需分别进行核心读者类群的识别及其借阅图书的数据分析，图书流通影响因素 H 类指数遴选结果如表 3 所示：

表 3 图书流通各影响因素 H 类指数筛选情况

H 类指数及相关性	一级类目－读者单位	一级类目－读者专业	一级类目－出版者	一级类目－出版年	一级类目－读者性别	一级类目－读者年级	一级类目－读者类别
H-A 指数	A	A	A	A	A	A	H-A
H-G 指数	A	A	A	A	A	A	H-G
A-G 指数	A	A	A	A	A	A	A-G

3.2 A 指数核心读者群分析

笔者基于 H 类指数的遴选，采用 A 指数分别分析一级类目与单位、学历、年级、性别、出版者、出版年和专业矩阵中各侧面核心读者群及其核心借阅藏书群、核心出版者群及核心出版年集群。

3.2.1 一级类目－单位 A 指数分析

表 4 显示：2014－2018 年分析了 23 个单位，总体 A 指数平均值为 324.6，总体均值以上的有经济管理学院等前 15 个单位，为我校图书馆的核心读者群；排在 16－23 的葡萄酒学院等 8 个单位属于非核心读者群。

表 4 图书流通数据一级类目－单位 A 指数排序（2014－2018 年）

单位	平均 A 指数	单位	平均 A 指数	单位	平均 A 指数
1. 经济管理学院	688.8	9. 资源环境学院	404.5	17. 外语系	198.9
2. 水利与建筑工程学院	530.5	10. 生命科学学院	400.0	18. 创新实验学院	176.9
3. 风景园林艺术学院	461.0	11. 食品科学与工程学院	363.3	19. 化学与药学院	151.0
4. 人文学院	442.2	12. 农学院	363.1	20. 理学院	147.2
5. 园艺学院	421.9	13. 动物科技学院	363.0	21. 成人教育学院	86.2
6. 林学院	417.8	14. 植物保护学院	360.0	22. 水土保持研究所	79.8
7. 机械与电子工程学院	415.4	15. 动物医学院	331.8	23. 马克思主义学院	15.4
8. 信息工程学院	411.4	16. 葡萄酒学院	235.4		

进一步分析图 2 可知：2014－2018 年各单位读者群藏书借阅 A 指数略有波动总体均呈下降趋势；平均 A 指数为 324.6 以上的 15 个核心单位中经济管理学院读者群 A 指数最高，处于第一核心读者群；第二读者群是以水利与建筑工程学院和风景园林艺术学院；第三核心读者群分属于资源环境学院等 12 个学院；葡萄酒学院等 8 个学院为非核心读者群。

3.2.2 一级类目－学历年级 A 指数分析

表 5 显示：2014－2018 年图书借阅总量平均 A 指数为 539.8 以上的 2014－2016 级本科生为核心读者群，通过数据年度分析发现：图书借阅统计年前 1、2 年进校的本科生依次为核心读者群和次核心读者群，其 A 指数显著高于其进校年和毕业年，可知大二、大三本科生是图书馆纸质藏书的核心读者群；硕士生核心读者群为 2013－2017 级硕士生，即二年级硕士生为第一核心读者群，一年级和三年级硕士生交替分获第二、第三核心读者群；第一核心读者群 A 指数显著高于第二核心读者群，是我校图书馆重要读者群；博士生、博士后读者群 A 指数差别不大，说明博士生、博士后读者群各年级核心程度均衡。

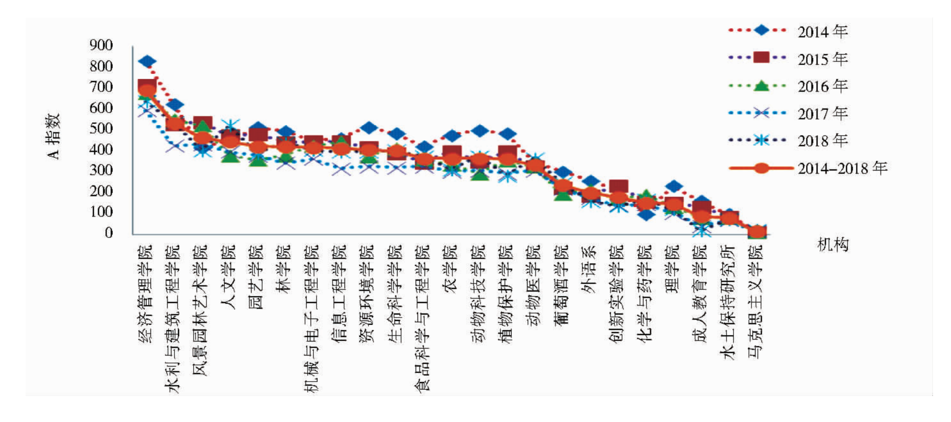


图2　2014－2018年各学院读者A指数年度分布

表5　2014－2018年一级类目－学历年级图书流通数据A指数分析

年级	本科生	硕士生	博士生、博士后
2010	108.1	2.1	8.9
2011	393.8	24.2	16.8
2012	660.8	78.6	25.1
2013	927.9	194.1	45.5
2014	1013.9	253.7	46.5
2015	758.8	231.8	44.1
2016	554.1	196.5	30.8
2017	357.3	166.5	27.5
2018	83.8	52.3	10.9
平均A指数	539.8	120.6	26.2
总平均A指数		215	

3.2.3　图书流通数据一级类目－类别A指数、H指数、G指数分析

一级类目－类别中的H指数、A指数和G指数两两相关性低，因此从H指数、A指数和G指数3个视角分析读者类别。表6显示，2014－2018年本科生和硕士生借阅H指数在平均值以上；A指数总体均值为1 557.5，本科生A指数为4 468.1，显著高于总体平均值，硕士生A指数为1 449.9，接近总体平均值，H指数、A指数筛选结果基本一致，而G指数总体平均值及其区分度受限，确定核心读者不够准确，因此本科生、硕士生为图书馆核心读者。

表6　2014－2018年图书流通数据一级类目－类别A、H、G指数分析

H类指数	H指数(平均H指数17.1)			A指数(平均A指数1 557.5)			G指数(平均G指数21.3)		
读者类别	本科生	硕士生	博士生，博士后	本科生	硕士生	博士生，博士后	本科生	硕士生	博士生，博士后
2014－2018年平均数	20.8	19.2	14.8	4 468.1	1 449.9	188.5	21.2	21.2	21.4

3.2.4　一级类目－性别A指数分析

2014－2015年男性读者A指数(5 449、3 340.2)显著高于女性读者(1 752.6、2 482.6)，随着时间的推移，女性读者的借阅A指数逐渐提高，2016－2018年女性读者A指数(2 975.2、3 098.6、4 277.0)逐渐高于男性读者(2 305.7、1 881.8、2 249.3)，并且这种A指数的差异呈扩大趋势。

3.3　一级类目－出版者A指数分析

表7可见，我校图书馆2014－2018年藏书涉及533家出版者，平均A指数(43.2)以上者共128家，这是我校图书馆藏书建设的重点出版社。其中排名第1的作家出版社A指数(891.6)，比排名第128的当代世界出版社A指数高约19.6倍，数据相差悬殊，因此对这128个出版社进行分类，将其分为核心区、相关区和非相关区。各个区的A指数之和相等时，其核心区、相关区、非相关区出版社数量呈1∶n∶n2，符合文献计量学中的布拉德福定律[17]。2014－2018年总体平均A指数按照此定律分为3个区，核心区出版者12家，相关区出版者32家，非相关区出版者84家。

3.4　一级类目－出版年A指数分析

2014－2018年我校图书馆共购买了1952－2018年出版的图书(含早年影印书籍)。由表8可见，其中2004－2017年出版图书A指数＞平均A指数(即核心出版年)，年度分析表明，借阅年前2年到前5年图书借阅A指数最高，是图书典藏、图书订购最重要的年分，借阅年向前1年倒推至前14年是需要着重考虑的藏书订购年和典藏年。

4　讨论

4.1　数据处理步骤设置

通过中图法一级类目与藏书流通数据7个影响因素之间相关性分析，确保了原始数据的可靠性；通过H

表 7 2014－2018 年图书流通数据一级类目－出版者 A 指数分区情况

核心区(A 指数)	作家出版社、广州出版社、北京十月文艺出版社、人民文学出版社、科学出版社、湖南文艺出版社、清华大学出版社、译林出版社、珠海出版社、人民邮电出版社、机械工业出版社、外语教学与研究出版社(12 家,A 指数 891.6－269.2)
相关区(A 指数)	长江文艺出版社、中国人民大学出版社、中国宇航出版社、河南文艺出版社、百花洲文艺出版社、上海译文出版社、化学工业出版社、中国水利水电出版社、中国石化出版社、高等教育出版社、江苏文艺出版社、花山文艺出版社、浙江文艺出版社、南海出版公司、大连理工大学出版社、天津人民出版社、上海文艺出版社、中国对外翻译出版有限公司、中国林业出版社、中信出版社、中国农业大学出版社、上海人民美术出版社、文汇出版社、北京大学出版社、重庆出版集团、电子工业出版社、航空工业出版社、时代文艺出版社、上海世界图书出版公司、中国农业出版社、外文出版社、海豚出版社(32 家,A 指数 268.4－124.3)
非相关区(A 指数)	江苏凤凰文艺出版社、北京联合出版公司、上海三联书店、陕西师范大学出版社、崇文书局、中国华侨出版社、中华书局、广西师范大学出版社、世界图书出版公司西安公司、中国青年出版社、中国建筑工业出版社、朝华出版社、华中科技大学出版社、中国国际广播出版社、华东理工大学出版社、上海人民出版社、国际文化出版公司、花城出版社、北京理工大学出版社、商务印书馆、中国纺织出版社、群言出版社、北京语言大学出版社、漓江出版社、安徽文艺出版社、九州出版社、青岛出版社、接力出版社、同济大学出版社、上海外语教育出版社、中国法制出版社、吉林美术出版社、万卷出版公司、上海文化出版社、新星出版社、长江出版社、吉林出版集团有限责任公司、西安交通大学出版社、南京大学出版社、西北工业大学出版社、贵州人民出版社、北京航空航天大学出版社、东南大学出版社、新世界出版社、北岳文艺出版社、二十一世纪出版社、江苏人民出版社、四川文艺出版社、国防工业出版社、辽宁美术出版社、线装书局、春风文艺出版社、东北大学出版社、天津大学出版社、中国轻工业出版社、上海交通大学出版社、生活·读书·新知三联书店、北方文艺出版社、北京邮电大学出版社、首都师范大学出版社、哈尔滨出版社、武汉出版社、东华大学出版社、西北农林科技大学出版社、北京燕山出版社凤凰出版社、中国政法大学出版社、中国铁道出版社、光明日报出版社、金城出版社、中国友谊出版公司、浙江教育出版社、中央编译出版社、百花文艺出版社、人民日报出版社、辽宁人民出版社、中国电力出版社、上海社会科学院出版社中国书籍出版社、武汉大学出版社、江苏教育出版社、浙江工商大学出版社、石油工业出版社、当代世界出版社(84 家,A 指数 43.3－121.2)

表 8 图书流通数据一级类目－出版年 A 指数分析

出版年	2012	2013	2011	2014	2010	2015	2009	2016	2008	2007	2017	2006	2005	2004	平均 A 指数
A 指数	1015.2	849.9	794.8	769.9	539.8	478.1	474.6	296.4	269.7	240.7	210.6	175.7	156.6	138.7	131.9

指数、A 指数和 G 指数相关性、变异性和适应性分析实证了 A 指数适用于分析高校图书馆图书流通数据；藏书流通数据 H 类指数相关性图谱说明了 H 指数和 A 指数的适用范围，为高校图书馆推广使用 A 指数分析藏书流通数据，加强藏书管理提供更多参考依据。

4.2 方法延伸与推广

本文基于我校图书馆藏书流通数据实证了平均 A 指数法分析的合理性，多视角计算核心读者群、核心出版者和核心出版年。该方法延伸与推广为：可以进行矩阵转置，通过各分类口径核心读者群映射核心藏书群；可以通过 A 指数直接分析核心藏书群，对结果比较分析可更加合理地确立核心藏书群；可以根据具体情况比较分析 A 指数的平均数、众数和中位数，从中选用最适合藏书流通数据的 A 指数；可针对 7 个影响因素分别进行更加深入的研究；可基于总体分析进行 7 个方面流通数据 A 指数的年度、季度和月度分析，以便观察流通数据的发展态势。

4.3 A 指数分析藏书影响因素的讨论

(1)我校 15 个核心读者学院均是建院时间长、综合实力较强的院系，这与核心借阅单位建院时间长、书香底蕴积淀及其对学生读书习惯培养有关，同时与一些学院(如经管学院、人文学院)资料性较强有关；一些非核心读者群所属学院或是近十几年建立，或是其任务是在面向全校授课和相关基础性研究，如外语系、理学院和马克思学院等。

(2)本科 2－3 年级学生借阅 A 指数显著高于进校年和毕业年，是我校图书馆纸质藏书核心读者群，这与他们进校第一年基础课程多，专业学习内容不够明确有关，2－3 年级后随着专业学习的深入逐步明确了个人需要和兴趣，进入高效学习阶段，阅读量显著增加，而在第四学年借阅 A 指数的降低与他们面临考研精准学习和找工作有关。硕士生二年级借阅 A 指数显著高于第一年和毕业年，说明硕士二年级已进入稳定、大量阅读的科研阶段，而进校年和毕业年借阅 A 指数较低，这与第一年从学习知识向创造知识的转型速度有关，与毕业年的重点任务转移(系统整理和分析实验数据、撰写毕业论文、找工作)有关，当本科阶段提前进入实验研究时，其硕士阶段稳定、大量阅读的时间也会提前；博士、博士后读者群 A 指数及其差异不大，主要以期刊为主实时关注科学前沿。

(3)当前许多高校图书馆藏书库容有限，因此核心出版社的遴选及其采用布拉德福定律分区的方法作为观察、评价、调整藏书采购格局、合理设置藏书典藏年的参考依据。

(4)据调研，我校近 5 年每年本科招生(5 300 人左右)男女比例范围 52.7%－48.2%，平均比例约 1∶1，硕士生(2 950 人左右)男女比例约 4∶6，而女性读者图书阅读 A 指数逐渐高于男性读者，并且这种差异呈

逐年扩大态势。

4.4 A 指数读者分类口径和规模的设置

本文进行的一级类目与读者单位、读者的学历、年级、出版者、出版年、性别和读者类别的 A 指数分析数据分别能够相互验证和支持，这从实际应用视角证明了 A 指数适用于藏书流通数据分析。而一级类目－单位与一级类目－学位授予专业 A 指数分析结果不能相互支持，这与学位授予专业所含学院数量差异显著，读者人数的非均衡现象有关，因此图书流通数据 A 指数分析时，应确定合理的读者分类口径。

5 结论

5.1 平均 A 指数分析藏书流通数据方法科学、可靠

基于 H 类指数遴选 A 指数分析流通数据解决了描述统计中数据之间缺乏可比性的问题和 H 指数极值钝化现象，平均 A 指数遴选核心读者群、核心出版社及核心出版年，其结果能够相互验证，从方法到实际应用均证实平均 A 指数分析图书流通数据合理、科学、可靠。

5.2 H 指数、A 指数、G 指数提供信息具有一致性

A 指数适合大数据分析，除一级类目－读者类别外，其余 6 个影响因素的 H 指数、A 指数、G 指数之间均呈高度相关性或显著相关，证实所提供的信息具有一致性；其中一级类目－年级 A 指数、G 指数，一级类目－出版者 H 指数、A 指数相关性呈有条件（A 指数 > ＝400，H 指数 > ＝10）显著相关，因此 A 指数适合一定量大数据分析。

5.3 多视角科学观察

选择 A 指数对 2014－2018 年前 20% 藏书流通数据从单位、专业、性别、学历年级等方面能够遴选我校图书馆核心读者群、核心出版年和核心出版者；当读者类别 H 类指数间相关性较弱时，可从 H 类指数诸视角联合分析，以便实现流通数据科学管理。

5.4 布拉德福定律适用于藏书出版者分类

根据 A 指数之和相等原则将平均 A 指数以上的 128 出版者分为 3 个区域（12:32:84），符合布拉德福定律。实证了布拉德福定律既适用于特定主题期刊管理，也适用于高校图书馆图书核心出版者、相关出版者及非相关出版者的确认，实现藏书优化管理。

总体来看，本文对 2014－2018 年前 20% 的流通数据 A 指数分析证明：①本科生、硕士生是图书借阅的核心读者群，并且本科生读者在校大二大三两年、硕士生在校第二年是图书馆纸质藏书最为核心的读者群。②平均 A 指数对核心读者单位的遴选说明我校图书馆纸质藏书重点考虑了我校专业历史悠久、品牌较大的学院的读者需求。③从性别借阅 A 指数的转移和趋势看，关注和调研性别阅读偏好，加强藏书建设和导读工作尤其重要。④我校图书馆核心出版年的确立为高校图书典藏和订阅提供参考依据。

参考文献：

[1] RONALD R，刘俊婉，马建华. Hirsc h 指数研究的新进展[J]. 科学观察，2006(4)：23－25.
[2] 周志峰. H 指数应用与图书馆借阅数据分析的探索[J]. 图书馆建设，2009(11)：83－89.
[3] 金碧辉，ROUSSEAU R. R 指数、AR 指数：H 指数功能扩展的补充指标[J]. 科学观察，2007(3)：1－8.
[4] HIRSC H. An index to quantify an individual's scientific research output [J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(15)：16569－16572.
[5] EGGHE L. How to improve of the H-index[J]. The scientist, 2006，2(1)：8－9.
[6] EGGHE L. An improvement of the H-indexthe G-index[J]. ISSI newletter，2007, 2(1)：8－9.
[7] 刘玉仙，ROUSSEAU R. Hirsch 型索引与图书馆管理：以同济大学图书馆为例 [C]//ISSI 2007 论文集. 马德里：CINDOC-CSIC, 2007：514－522.
[8] 梁晓虹. 基于借阅流通数据的高校图书馆纸质馆藏资源利用研究[D]. 济南：山东大学，2018.
[9] 周春雷. 基于 H 指数的学术授信评价研究[D]. 武汉：武汉大学，2010.
[10] 孔超，丁璇. g 指数在高校图书馆借阅数据分析中的探索[J]. 现代情报，2014，34(12)：126－129，133.
[11] 钱玲飞，汪荣. 基于 H 指数的 OPAC 数据分析及应用——以南京航空航天大学图书馆为例[J]. 大学图书馆学报，2012，30(2)：61－64.
[12] 张晓阳，方友亮，宋新平. 图书引用对学术绩效 h 指数的影响——以图书情报学领域为例[J]. 科学学研究，2014，32(2)：189－194.
[13] 李雅，姚立会，郭虹. 由植保学博士论文引文特征看我馆期刊资料保障率[J]. 农业图书情报学刊，2006(2)：156－157，160.
[14] 陈俊科，李红，李捷，等. 基于变异系数法的新疆畜牧业发展综合效益评价[J]. 黑龙江畜牧兽医，2015(18)：17－20.
[15] 黄良文，陈仁恩. 统计学原理[M]. 北京：中央广播电视大学出版社，1996.
[16] 蒲筱哥，刘礼明. 数据库绩效的变异系数与灰色关联度分析及实证研究[J]. 图书情报工作，2014，58(14)：71－78.
[17] 祁延莉，刘西琴. 核心专利识别方法研究[J]. 情报理论与实践，2016，39(11)：5－9.
[18] 刘莲花. 基于距离综合评价法的数学核心期刊评价[J]. 中国科技期刊研究，2015，26(8)：871－874.

作者贡献说明：
净玲娣：负责统计方法实施、论文结构梳理和论文撰写、修改与完善；
王立宏：负责原始数据的提取、规范化清理和 A 指数函数批计算；
王芹：负责 R 软件对数据相关分析的批计算、论文修改；
李雅：负责文章整体思路、结构的规划、图表制作、论文的整体把关与修改论文。

Selection of H-family Indices for Book Circulation Data Analysis
——Taking the Library of Northwest Agriculture and Forestry University as an Example

Jing Lingdi Wang Lihong Wang Qin Li Ya
Library of Northwest Agriculture & Forestry University, Yangling 712100

Abstract: [Purpose/significance] To improve the management of book procurement and circulation, and to realize the standardization management of timely monitoring, and evaluation of the collection acquisition, the total volume and structure of the collection, etc. [Method/process] We compared the correlation, variability and adaptability of three H-family indices H-index, A-index and G-index. Then we choosed H-index as present of suitability for the evaluation of book circulation. And we used H-index to analyze the top 20% of NWAFU library's book circulation data from 2014 to 2018. [Result/conclusion] The result of analysis shows that A-index is a more comprehensive and precise indicator with ideal performance for identifying the core readers, core publishers and core publication years when the statistical relation between first category of Chinese Library Classification together with reader's gender, readers' major, readers' grade, readers' department, book's publisher and book's publication year are concerned, while it is able to select core readers, core publishers and core years more comprehensively and accurately. Besides, Bradford law can not only be applied to periodical management but also be applied to book management. This paper can provide a reference a basis for exploring the use of H-index in analyzing book circulation data.

Keywords: H-index average A-index coefficient of variation correlation book circulation data